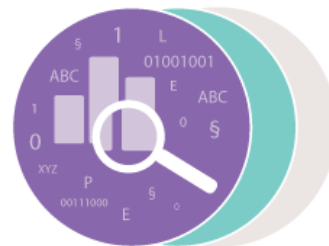




# FutureTDM

Explore . Analyse . Improve



REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

## Deliverable 6.4

### FutureTDM policy frameworks, roadmaps, and practitioner guidelines (Additional Report)

## Project

Acronym:	<b>FutureTDM</b>
Title:	Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach
Coordinator:	SYNYO GmbH
Reference:	665940
Type:	Collaborative project
Programme:	HORIZON 2020
Theme:	GARRI-3-2014 - Scientific Information in the Digital Age: Text and Data Mining (TDM)
Start:	01. September, 2015
Duration:	24 months
Website:	<a href="http://www.futuretdm.eu/">http://www.futuretdm.eu/</a>
E-Mail:	<a href="mailto:office@futuretdm.eu">office@futuretdm.eu</a>
Consortium:	<b>SYNYO GmbH</b> , Research & Development Department, Austria, (SYNYO) <b>Stichting LIBER</b> , The Netherlands, (LIBER) <b>Open Knowledge</b> , UK, (OK/CM) <b>Radboud University</b> , Centre for Language Studies The Netherlands, (RU) <b>The British Library Board</b> , UK, (BL) <b>Universiteit van Amsterdam</b> , Inst. for Information Law, The Netherlands, (UVA) <b>Athena Research and Innovation Centre in Information, Communication and Knowledge Technologies</b> , Inst. for Language and Speech Processing, Greece, (ARC) <b>Ubiquity Press Limited</b> , UK, (UP) <b>Fundacja Projekt: Polska</b> , Poland, (FPP)

## Deliverable

Number:	<b>D6.4</b>
Title:	<b>FutureTDM policy frameworks, roadmaps, and practitioner guidelines</b>
Lead beneficiary:	<b>ARC</b>
Work package:	WP6: Empower: Collaborative knowledge base, open information hub, awareness raising
Dissemination level:	Public (PU)
Nature:	Other (O) Additional Report
Due date:	31.07.2017
Submission date:	31.07.2017
Authors:	<b>Kanella Pouli, ARC</b> <b>Burcu Akinci, SYNYO</b>
Contributors:	<b>Kiera McNeice, BL</b>
Review:	<b>Bernhard Jäger, SYNYO</b>

**Acknowledgement:** This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 665940.

**Disclaimer:** The content of this publication is the sole responsibility of the authors, and does not in any way represent the view of the European Commission or its services.

This report by FutureTDM Consortium members can be reused under the [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) license (<https://creativecommons.org/licenses/by/4.0/>).



**Table of Contents**

1 INTRODUCTION ..... 6

2 POLICY FRAMEWORK ..... 10

3 PRACTITIONER GUIDELINES..... 14

4 ROADMAP..... 19

5 CONCLUSION ..... 21

## Table of Contents

Figure 1: The FutureTDM home page .....	6
Figure 2: The TDM Projects page .....	7
Figure 3: The TDM Methods page .....	8
Figure 4: TDM Tools Detail Page .....	8
Figure 5: Instructions on how to read the policy framework.....	11
Figure 6: Barriers and the legend for color coding the stakeholders.....	11
Figure 7: The TDM value chain .....	12
Figure 8: Legend for coding the stakeholders .....	12
Figure 9: Recommendations concerning awareness and clarity in Technical and Infrastructure .....	13
Figure 10: Practitioner Guidelines .....	14
Figure 11: Legal Guidelines for TDM Practitioners.....	15
Figure 12: Do's and Don'ts in the Legal Guidelines section .....	15
Figure 13: Summary of key points in the Guidelines for Content Licences section .....	16
Figure 14: Data management benefits.....	16
Figure 15: Data management guidelines for researchers .....	17
Figure 16: Guidelines for supporting TDM at Universities .....	17
Figure 17: Benefits to universities.....	18
Figure 18: Roadmap .....	19
Figure 19: Roadmap Phase I: Content Availability .....	20
Figure 20: Feedback requested on the FTDM platform .....	21

# 1 INTRODUCTION

This report presents the outcomes of the FutureTDM project as concerns the **elaborated frameworks**, **policy priorities**, **roadmaps** and **practitioner guidelines** as they have been published on the FutureTDM platform via informative and insightful infographics and visualisations to appeal to a dynamic range of users and stakeholders.

The [FutureTDM platform](#) consists of two parts: the main dissemination platform namely **Open Information Hub** and the **Collaborative Knowledge Base** where the collection of knowledge and information resources on Text and Data Mining is displayed in a user-friendly structure with coherent and up-to-date view on the TDM landscape. The platform<sup>1</sup> was launched on April 1<sup>st</sup>, 2017 and has the following structure:

1. [Home page](#): it serves as an entry point, provides information on the project, and has the latest updates on events, newsletters, blog posts and videos (see Figure 1).

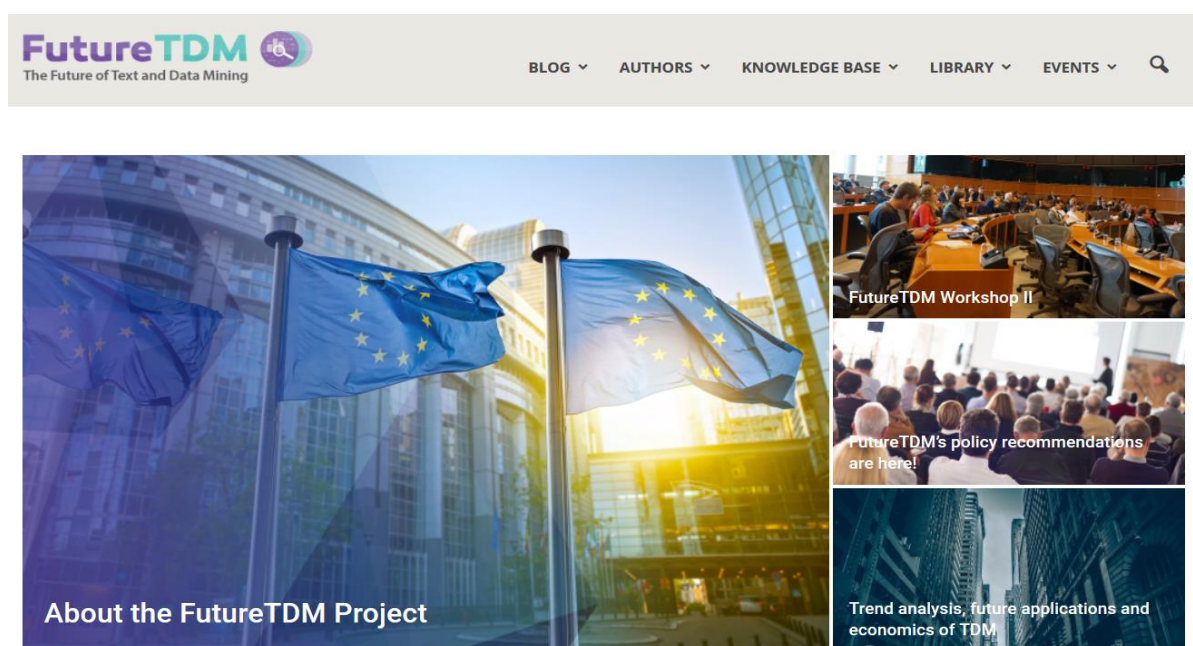
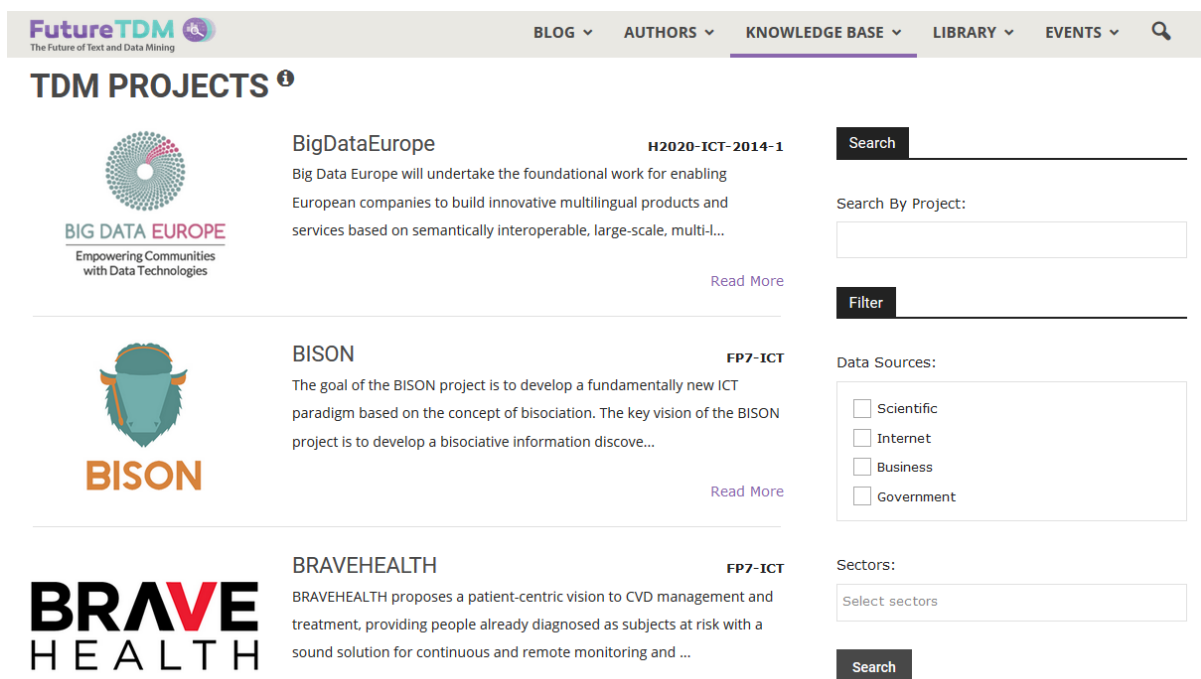


Figure 1: The FutureTDM home page

2. [Blog](#): it contains personal opinions of the authors on important TDM issues as well as some of the project outcomes under the categories listed below:
  - Community & Events
  - Education & Skills
  - Legal & Policies
  - Research & Insights

<sup>1</sup> The platform was created using standard open access software and designed with a flexible and scalable architecture to guarantee a long-term sustainable usage. It was launched eight months earlier than initially planned. It is described in detail in deliverables 6.1 Web-based FutureTDM Collaborative Knowledge Hub, and 6.3 Research projects directory and best practice library (Progress Report).

- Stories & Cases
  - Technologies & Tools
3. [Authors](#): this section contains lists of the organizations, the projects and the people (internal and external experts) who are involved in a variety of ways with the FutureTDM project.
  4. [Knowledge Base](#): the content of the Knowledge Base has been derived from various project deliverables<sup>2</sup>; it is organized in the following sections and hopefully serves as a reference point for researchers, practitioners and any stakeholder interested in the field.
    - [TDM Projects](#) is a directory of the most important TDM related projects (see Figure 2)




**Figure 2: The TDM Projects page**

- [Expert Navigator](#) is a directory of the organisations working with TDM
- [TDM Methods](#) is a directory of the commonly used methods for TDM (see Figure 3)
- [TDM Tools](#) is a directory of tools used for TDM,
- the [Policy Framework](#) section is dedicated to recommendations for TDM practices (for a detailed description see section Policy Framework),
- the [Practitioner Guidelines](#) section contains all the sets of guidelines to help address specific challenges to greater use of TDM technologies (for a detailed description see section Practitioner Guidelines ), and
- the [Roadmap](#) section outlines a plan to increase the uptake of TDM technologies in Europe (for a detailed description see section Roadmap).

<sup>2</sup> The Knowledge Base presents collections derived from WP4, specifically D4.2 Collections of relevant projects, (not publicly available), which encompasses individual persons or groups of people focusing on TDM (i.e. experts, organizations and project collections) as well as technologies and resources useful for current and future TDM practitioners (i.e. methods and tools lists).


## TDM METHODS <sup>1</sup>



### Artificial Intelligence

Artificial intelligence (AI) can be defined as an intelligence exhibited by machines. The term is frequently applied to The art of creating machines that perform functions that require intelligence...


[Read More](#)



### Association Rules

Association rules are rule-based machine learning methods for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases...

[Read More](#)



### Classification

Classification is a general process related to categorization, in which ideas and objects are recognized, differentiated, and understood. A classification system is an approach to accomplishing cla...

[Read More](#)

Search By Method:

**Filter**

Sectors:

**Search**

**HELP US IMPROVE OUR KNOWLEDGE BASE**

FutureTDM Knowledge Base is continuously collecting information related to TDM field in Europe. Let us know what you are looking for? Help us improve our content.

Your feedback is important for us and it will be fed into our Knowledge Base.

**Figure 3: The TDM Methods page**

In addition functionalities have been created to facilitate search in the directories which are dedicated to projects, experts, methods and tools (see Figure 2: The TDM Projects page Figure 2 and Figure 3): the user can search either by using a name or by filtering results according to data sources of usage/specialization, or even specific sectors and/or application fields of interest.


Below (See Figure 4) detail page of one of the TDM Tools, ContentMine, can be seen.

**FutureTDM**  
The Future of Text and Data Mining

BLOG ▾ AUTHORS ▾ KNOWLEDGE BASE ▾ LIBRARY ▾ EVENTS ▾

Home ▾ Tool Profile

## CONTENTMINE



ContentMine

Website: [www.contentmine.org](http://www.contentmine.org)

**Method Used:** Data Mining, Text Mining

**Tool Types:** Open Source, Web Application, Pipeline




**Data Sources:** Scientific

**ABOUT**

The software stack provides tools for getting papers from many online sources, normalising them, then processing them to lookup and/or search for key terms, phrases, patterns, statements, and more. The stream of facts is available in a standardized, easy-to-reuse way and that makes it share-able.


Note: The information is gathered through press information or publicly published About pages of the companies, unless stated otherwise.

**ADDITIONAL INFO**






[Share on Facebook](#)
[Tweet on Twitter](#)
[G+](#)
[in](#)


**LATEST POSTS**




Jan Borychalski July 6, 2017




June 29, 2017



June 2, 2017



May 17, 2017



May 15, 2017

**Figure 4: TDM Tools Detail Page**



5. The [Library](#) is the section dedicated to all the project produced material: the publicly available deliverables which can be found under the category “expert reports”, the flyers (available at “Dissemination outputs”) as well as all the Awareness Sheets.
6. [Events](#): this section is subdivided into the Knowledge Cafés<sup>3</sup>, Community Events and Event photos, providing information on the various events organized by FTDM or to which FTDM participated.

Besides the strictly informative nature of the platform on issues related to TDM, special care has been given to create divisions providing insights and solutions to TDM practitioners and stakeholders. These are the sections dedicated to the **policy framework**, the **guidelines** and the **roadmap** which are described in detail in the following chapters. As dictated by the Description of Action, special design and architecture has been used to make the visualization of the project results as explicit and comprehensive as possible. Different colours have been chosen for each of the aforementioned sections, each one being associated with the FTDM logo: **purple** for the policy framework, **light blue** for the guidelines and **light green** for the roadmap.

---

<sup>3</sup> Throughout the first half of 2016, FutureTDM had been running a series of Knowledge Cafés across Europe, the results of which are presented in **D2.3 Report on stakeholder mobilisation and perceptions**, and summarized in this section of the platform.

## 2 POLICY FRAMEWORK

The ongoing dialogue with the various TDM groups throughout the project<sup>4</sup>, as well as the close collaboration with the advisory board of the consortium, ensured various stakeholder perspectives and highlighted the problems encountered by the TDM community. The barriers encountered were grouped in four broad categories:

- a) Legal Rules & Policies
- b) Education, Skills & Education
- c) Economy & Incentives
- d) Technical & Infrastructure

To overcome these barriers a hierarchical model of principles and recommendations has been created and presented in the [Policy Framework](#) section of the Knowledge Base on the platform.

A guide on how to read the recommendations is provided by the drop-down heading "HOW TO READ" (see Figure 5). The overall principles underlying the barriers in all categories are (as explained in [D5.1 FutureTDM Policy Framework](#)):


- **Uncertainty:** this category includes uncertainties as to how, why and if TDM can be carried out, as well as the lack of awareness of different aspects of TDM,
- **Fragmentation:** this refers to the fragmentation in the TDM landscape, which prevents TDM from being carried out across e.g. national borders, scientific domains, companies or fields of expertise, and
- **Restrictiveness:** the last category refers to direct limitations to possibilities to carry out TDM, either by restrictive laws, lack of expertise, limited (financial) resources, etc.

The recommendations created by FTDM experts are based on the principles of:


- **Awareness & Clarity**, which aims to solve the barriers related to uncertainty, but also the lack of awareness,
- **TDM Without Boundaries**, which aims to solve the issues of a fragmented TDM landscape, and
- **Equitable Access**, which aims to deal with restrictions -either legal, practical, economic or technical- to TDM.

---


<sup>4</sup> The outcomes of consultations with stakeholders are presented in deliverables **D2.3 Report on stakeholder mobilisation and perceptions**, [D2.4 Workshop summary report 1](#), [D2.5 Workshop Summary report 2](#), [D4.3 Compendium of Best Practices and-Methodologies](#), [D4.5 Compendium of Best Practices and Methodologies](#) (the deliverables which are publicly available can be downloaded from the [Open Knowledge Library](#)). They are licensed under a [Creative Commons Attribution 4.0 International License](#) (CC-BY) and appropriate attribution to the FutureTDM project is mandatory when sharing and adopting the provided material.

**HOW TO READ** 


For each area, we first provide a short **recap of the barriers** that we have established. Next, the recommendations will follow, which aim to overcome these barriers. For this purpose, we have formulated **three fundamental principles** that stakeholders should take into account when addressing barriers:

 **Awareness & Clarity**

The benefits of TDM should be known, and it should be clear what they are, and if they can, and how they can, be achieved.

 **TDM Without Boundaries**

Within Europe, TDM should not be subject to different national rules, skills and knowledge should flow across domains and sectors, and standardisation and interoperability should be achieved for data, software and infrastructure.

 **Equitable Access**

There should be access to sources, tools, infrastructure and money to enable TDM in academia, industry and other types of organisations. However, legitimate opposing interests should be taken into account as well.

Every fundamental principle has resulted in **subprinciples**, that will guide more specific (groups of) stakeholders, followed by more concrete **recommendations** aimed at certain stakeholders. Please hover with your mouse on footnotes or icons for additional information, where applicable.

**Figure 5: Instructions on how to read the policy framework**

The platform offers the possibility to navigate through the four categories, with a recap of barriers for each one preceding recommendations (see Figure 6).

LEGAL RULES AND POLICIES
EDUCATION, SKILLS AND AWARENESS
ECONOMY AND INCENTIVES
TECHNICAL AND INFRASTRUCTURE

**RECAP OF BARRIERS**

**Copyright & database law**

- Uncertain: scope of exceptions often unclear
- Fragmented: different national scope of exceptions
- Restrictive: broad reproduction right and narrow exceptions





**Data protection law**

- Uncertain: how to comply
- Fragmented: different national interpretations, bodies and policies
- Restrictive: TDM of personal data potentially unlawful

**Documentation**

- Uncertain: not always clear whether TDM permitted
- Fragmented: TDM practitioners have to deal with many different licence conditions
- Restrictive: TDM often not permitted under non-OA licences

**LEGEND FOR COLOUR CODING THE STAKEHOLDERS**

-  Stakeholders' action in Content creation
-  Stakeholders' action in Content dissemination
-  Stakeholders' action in Text and data mining
-  Stakeholders' action in Knowledge utilisation

**Figure 6: Barriers and the legend for color coding the stakeholders**

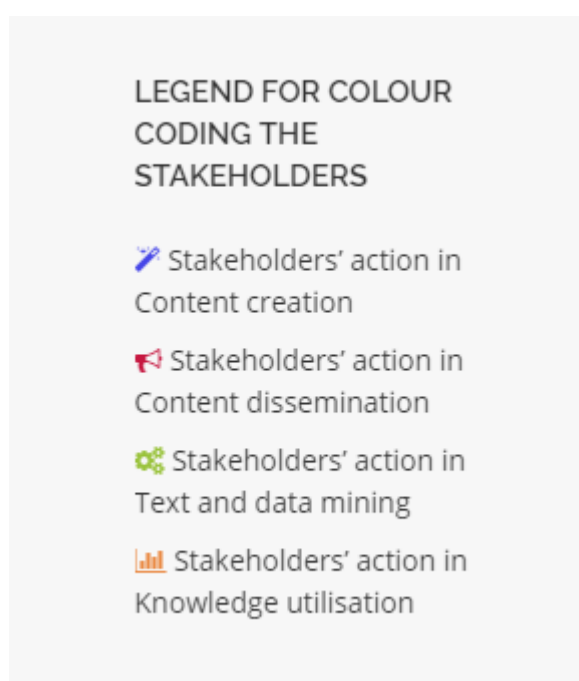
All recommendations for actions are accompanied by symbols which are explained by a legend appearing on the top right of each category section (see ). This legend makes reference to the steps of the TDM process, known as the TDM value chain<sup>5</sup> (see Figure 7) where different groups of TDM stakeholders are involved.

<sup>5</sup> For more information please see [D5.1 FutureTDM Policy Framework](#).



**Figure 7: The TDM value chain**

Therefore, in any given category (Figure 9 showcases an example from **Technical and Infrastructure**) first the principle under consideration is introduced (here it is **AWARENESS AND CLARITY**) and then the existing subprinciples are highlighted, each one in a purple box: Subprinciple 1 (Clear documentation) and Subprinciple 2 (Consistency and Completeness of datasets). After each subprinciple is briefly explained, the appropriate categories of stakeholders are mentioned with a symbol (see Figure 7) representing the area where the action should be made: content creation, content dissemination, text and data mining or knowledge utilization (this area is equivalent to the value creation step of the TDM value chain).



**Figure 8: Legend for coding the stakeholders**

In the case of subprinciple 1, it is developers who are responsible for writing clear documentation (and updating it regularly) and this affects the actual Text and Data Mining phase; on the other hand creators of datasets and metadata curators, should take action in both the stages of actual TDM and Knowledge Utilization.

The use of symbols and colours, as well as the common idea on which the information in all categories is organized and presented, facilitates the understanding of complex issues and problems which are

intertwined and which involve the active participation of multiple groups of stakeholders to overcome them.

LEGAL RULES AND POLICIES
EDUCATION, SKILLS AND AWARENESS
ECONOMY AND INCENTIVES
TECHNICAL AND INFRASTRUCTURE

### RECAP OF BARRIERS

**Copyright & database law**

- Uncertain: scope of exceptions often unclear
- Fragmented: different national scope of exceptions
- Restrictive: broad reproduction right and narrow exceptions

**Data protection law**

- Uncertain: how to comply
- Fragmented: different national interpretations, bodies and policies
- Restrictive: TDM of personal data potentially unlawful

**Documentation**

- Uncertain: not always clear whether TDM permitted
- Fragmented: TDM practitioners have to deal with many different licence conditions
- Restrictive: TDM often not permitted under non-OA licences

### LEGEND FOR COLOUR CODING THE STAKEHOLDERS

- Stakeholders' action in Content creation
- Stakeholders' action in Content dissemination
- Stakeholders' action in Text and data mining
- Stakeholders' action in Knowledge utilisation

### AWARENESS AND CLARITY

**Subprinciple 1 – Make clear rules on copyright, database and data protection law**

@Lawmaker

Minimise borderline cases where it is unclear whether copyright or related laws apply, by clearly defining terms and concepts (e.g. research organisation, text and data mining or scientific research purposes) in law-making »

@Policy makers

Provide explanatory documentation to accompany relevant laws, that TDM practitioners can refer to for guidance »

**Subprinciple 2 – Create guidelines on the law: what is permitted and what not?**

@FutureTDM

Provide practitioner guidelines on how the law works, what the rules are and how to best comply »

@European Data Protection Board & national data protection authorities

- Provide general guidelines on TDM to help practitioners comply with data protection law and other laws relating to information privacy and confidentiality
- Offer certification of data research and/or self-regulation and/or codes of conduct concerning TDM research or activities dealing with personal data, to provide certainty for practitioners
- Create guidelines on what is to be understood as archiving purposes and historical, statistical and scientific purposes to help practitioners understand when these apply »

@Professional associations representing 'personal-data intensive' companies and research institutes

Draft self-regulation or codes of conduct ensuring compliance, in particular with information privacy and data protection rules, for companies and research institutes to follow »

@Research institutes & (start-up) companies

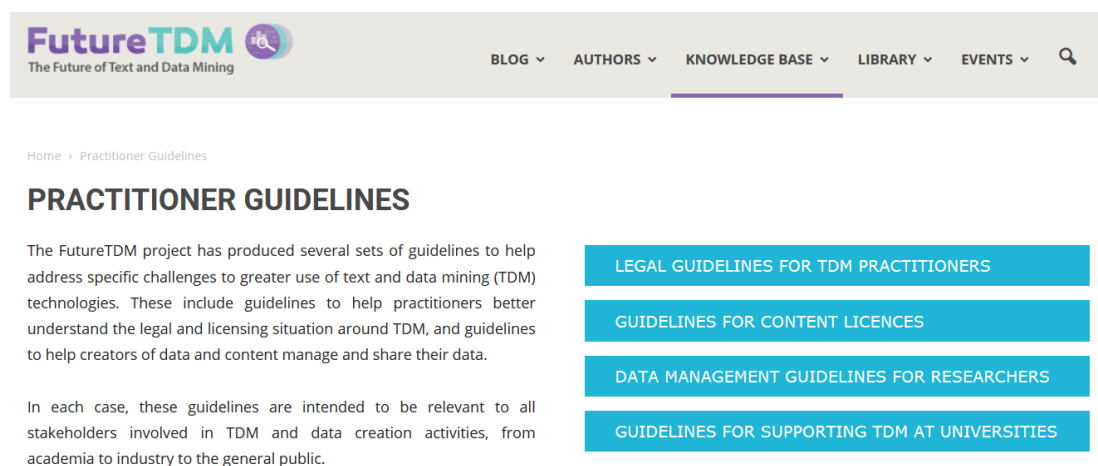
Implement ethical and legal review practices for large research projects dealing with big data »

**Figure 9: Recommendations concerning awareness and clarity in Technical and Infrastructure**

### 3 PRACTITIONER GUIDELINES

The Knowledge Base section devoted to the [practitioner guidelines](#) has been created to help address specific challenges to greater use of text and data mining (TDM) technologies. The FutureTDM project has produced several sets of guidelines<sup>6</sup> which are grouped in four categories (see Figure 10) addressing mainly, but not exclusively, data providers and users as anyone else, from academia to industry to the general public, who is interested in TDM. In specific one can find guidelines for:

- TDM practitioners on legal issues
- TDM practitioners on issues related to licences and licencing
- Researchers on Data Management
- Universities on curricula and education issues



**Figure 10: Practitioner Guidelines**

The first set of guidelines is entitled [Legal Guidelines for TDM Practitioners](#) and is based on the realization that the legal landscape around TDM is more complicated than TDM practitioners may believe. For this reason, it provides an overview of the legal landscape around TDM (e.g. what is public domain information, what is personal data etc.) so that legal risk is minimized and practitioners are protected. These guidelines are not intended to be comprehensive legal advice though; they, rather, aim to give TDM practitioners a foundational overview of relevant legal considerations and to help understand when it might be necessary to seek expert legal advice. In order to make it more comprehensible, information is presented either as a step to step plan, as for example when approaching content for mining, or broken down to questions and answers (see Figure 11).

<sup>6</sup> For a detailed description of the guidelines, please see D5.3 FutureTDM practitioner guidelines, available at [http://www.futuretdm.eu/wp-content/uploads/FutureTDM\\_D5.3-FutureTDM-practitioner-guidelines.pdf](http://www.futuretdm.eu/wp-content/uploads/FutureTDM_D5.3-FutureTDM-practitioner-guidelines.pdf)  
© 2017 FutureTDM | Horizon 2020 | GARRI-3-2014 | 665940

⌵ Do I need to search for the rights holder?

**Step-by-step plan to minimise risk**

To minimise risks, we advise you work through the following steps.

⌵ Step 1: Is it protected?

⌵ Step 2: What am I going to do with it?

⌵ Step 3: For what purpose?

⌵ Step 4: Do I have or need a licence?

⌵ Step 5: Do I need further legal assistance?

#### Public Domain?

**Copyright** lasts 70 years after the death of the author. Historical sources may be out of copyright. **Neighbouring rights** last 50 years after first publication, or 70 years in the case of phonograms. **Database rights** last 15 years after publication. If a database is substantially modified, this term starts again counting from the day the modified version is published.

#### Retrieval from databases

If you retrieve information from a database, this will not infringe any database rights if you only retrieve an insubstantial part. Retrieving substantial parts – at once or bit by bit – of the database as a whole does affect the rights holders' exclusive rights.

**Figure 11: Legal Guidelines for TDM Practitioners**

There is also an easily memorizable distinction of DO's and DON'Ts with lists of the permitted vs non-accepted TDM policies and practices (see Figure 12) as well as a link at the bottom of the page to the relevant deliverable for those who wish to delve into more detail.

⌵ Do's

- Establish if you will use or mine personal data and whether it also includes sensitive data
- Assign a Data Protection Officer if TDM is one your organisation's core activities, or if your organisation does TDM on a regular basis
- Impact Assessment (IA): establish what data you will use for what purposes, and who will have access to the data within and outside your organisation, and whether your use of personal data brings any high risks
- Check whether you have the legal grounds to collect and/or use the personal data
- Privacy by design: based on your IA, design your whole TDM project in a way that guarantees that you can safely and adequately use the personal data
- Look into sector-specific regulation, or self-regulation and codes of conduct within your domain, which may provide you more guidance and certainty on what you can do
- Anonymise data, so you are not dealing with personal data any more. Note that if you pseudonymise personal data, this is will still be personal data if the use of additional information enables you to attribute the data to a natural person.

⌵ Don'ts

For more information see our [Deliverable 5.3 FutureTDM Practitioner Guidelines](#)

#### Examples of personal data

• Name, age, gender • Home address • Phone number • Personal email • IP address • Bank account data • Passport data • Genetic data • Health data • Criminal records

#### Consent

Within the context of data protection law, consent by the data subject must be: • *Unambiguous*: no doubt may exist • *Informed*: all relevant information must be given to give informed consent • Registered properly, in able to prove and review the consent from each individual afterwards

#### Sensitive data

• racial or ethnic origin • political opinions • religious or philosophical beliefs • or trade union membership • genetic data • biometric data for the purpose of uniquely identifying a natural person • data concerning health • data concerning a natural person's sex life or sexual orientation

**Figure 12: Do's and Don'ts in the Legal Guidelines section**

The following set of guidelines is entitled [Guidelines for Content Licences](#) and is dedicated to various intellectual property (IP) laws that can be relevant to the use of content for TDM activities. It follows the same structure with questions and answers described earlier and in addition it offers a summary of the most important key points to be remembered (see Figure 13).



### Summary of key points

Until and unless the EU and its member states adopt consistent exceptions to intellectual property rights for the purposes of TDM, licences remain a key consideration for anyone planning to carry out TDM. Some key points to remember are:

- There are several kinds of legal restrictions that apply to TDM, beyond licensing; check section 2 to make sure you understand these as well.
- Even when an exception to IP rights applies, licence terms may affect your ability to carry out TDM; make sure you understand any relevant terms in your licence.
- If it is not possible to find a licence or identify the rights holder for a given piece of content, TDM may not be lawful; consider carefully whether you need expert advice on risk.
- If you negotiate licences on behalf of an institution, you play a key role in enabling those you represent to carry out TDM; please talk to your researchers, make sure you understand their needs, and consider whether licences are appropriate and reasonable for all parties.

For more information see our [Deliverable 5.3 FutureTDM Practitioner Guidelines](#)

**Figure 13: Summary of key points in the Guidelines for Content Licences section**

Next, there is a set of guidelines aiming primarily at academic researchers who collect, create, store and share data ([Data Management Guidelines for Researchers](#)) in order to help them make sure their data is reusable. This subsection clarifies what data management is and reveals all the benefits deriving from it.

### What are the benefits of data management?

#### ✎ For researchers (data providers)

Imagine a common scenario: a researcher has produced or collected data for their research, which need to be submitted to their organisation's repository, or the organisation that funded their research. If their organisation has an efficient Data Management Plan in place, the DMP's guidelines can help the researcher and their data to benefit from:

- **Discoverability:** When the data is registered in the organisation's inventory, catalogue, or repository, they become visible and discoverable by others.
- **Documentation:** When the data adhere to common standards for documentation, including metadata descriptions, they become valuable not only to human users, but to machine processes as well.
- **Security:** When the data are securely stored in the organisation's platform (which could be a repository or other type of organised storage facility), they are safeguarded and the risk of data loss is minimised.

**Figure 14: Data management benefits**

This subsection addresses specific questions, such as how to identify a repository in an appropriate domain for data storage, or guidelines on how to secure validation and quality of data and metadata



(see Figure 15). All the guidelines have been created in order to make data by researchers as valuable for reuse as possible.

**Data management guidelines for researchers**

If you are a researcher who has generated or collected data, how can you make sure your data is genuinely useful and re-usable when you deposit it in a repository? This section provides a set of guidelines to help you make your data as valuable as possible for future re-use.

- ✎ Identifying a repository in an appropriate domain
- ✎ Understanding metadata requirements
- ✎ Validation and quality assurance of data and metadata

Data quality must be defined in terms of a particular user and use case; a dataset might be perfect for one user's use case, but not so good for another. For example, a dataset from a medical database might be appropriate for a medical researcher who works on diabetes, but quite useless to a political scientist searching for patterns in protest movements.

**Figure 15: Data management guidelines for researchers**

Last, there is subsection devoted to Universities ([Guidelines for Supporting TDM at Universities](#)) as their role has been identified of crucial importance for the TDM uptake. Here the focus is on the necessary skills and education programmes to provide future users with them (see Figure 16).

**Skills and education**

Although not an explicit part of the TDM value chain, all aspects of the TDM landscape rely on stakeholders and practitioners having access to the necessary skills, education and support to carry out TDM projects. Universities play a crucial role in supplying the necessary skills, education and awareness that underpin all other aspects of the TDM value chain.

- ✎ The role of universities
- ✎ Benefits to universities

**Challenges**

Through interviews, workshops, and other consultations with stakeholders, the FutureTDM project identified several significant barriers that hinder greater uptake of TDM in universities. These are discussed in depth in the [FutureTDM report on policies and barriers to TDM in Europe](#), but the major barriers are summarised below.

- ✎ Lack of awareness
- ✎ Fragmentation of resources
- ✎ Skills gaps
- ✎ Lack of resources

**Content Creation**

Universities and their researchers generate large amounts of research data and other content that are potentially valuable to TDM practitioners. Making sure this content is managed and shared according to best practices offers benefits for universities both as content creators and owners, as well as potential re-users of others' content.

**Skills and Education**

All aspects of the TDM landscape rely on stakeholders and practitioners having access to the necessary skills, education and support to carry out TDM projects. Universities play a crucial role in supplying the necessary skills, education and awareness that underpin all other aspects of the TDM value chain.

**Lack of awareness**

Even though TDM technologies began emerging in the 1990s, it is still seen by many as a new, or even a "niche" field. Particularly outside of traditionally data-driven disciplines, there can be less awareness of or interest in TDM – despite its vast potential for

**Figure 16: Guidelines for supporting TDM at Universities**

Again, the possible benefits from university engagement with TDM is highlighted (see Figure 17). As explicitly stated "Data is the new IT" and Universities should be prepared to provide researchers with the necessary tools to make the best use of it.

### Universities as key stakeholders

It would be difficult to exaggerate the potential impact universities can have on the uptake of TDM technologies. In FutureTDM's policy framework and recommendations, we identified universities as influential stakeholders in all stages of the TDM value chain.<sup>2</sup>

- ✧ The TDM value chain
- ✧ The role of universities
- ✧ Benefits to universities

The principal reason for universities to invest resources in supporting TDM is that data science and analysis is fast becoming fundamental to all areas of research and education. Students need to understand data and how to manipulate it in order to be comfortable and confident in the modern world. In the words of one professor at a leading university, "Data science is the new IT."

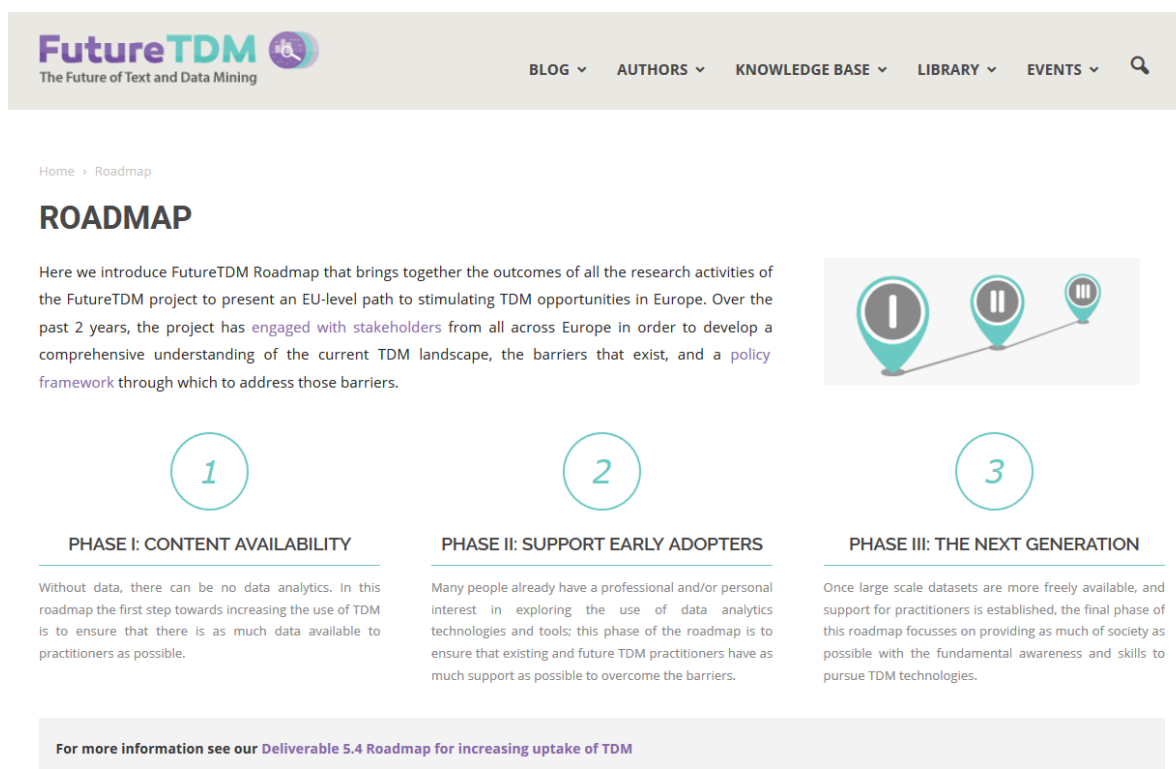
Given the exponential rate of growth of research outputs, the next generation of researchers will need to know how to employ data analysis techniques purely to keep up with the literature. And data analytics are already finding applications in what may seem to be unlikely fields. In the fashion industry, for example, business is largely conducted online – which means students need a sensitivity and awareness of the value that data analytics, including TDM, can bring to their business models.

**Figure 17: Benefits to universities**

## 4 ROADMAP

As mentioned in the previous chapters, over the past 22 months, the project has engaged with stakeholders from all across Europe<sup>7</sup> in order to develop a comprehensive understanding of the current TDM landscape, the barriers that exist, and a policy framework through which to address those barriers<sup>8</sup>. FutureTDM has created a roadmap that brings together the outcomes of all the research and mobilization activities of the FutureTDM project to present an EU-level path to stimulating TDM opportunities in Europe. This [roadmap](#) has its own dedicated section in the platform (see Figure 18): a plan is outlined to increase the uptake of TDM technologies in Europe, focusing on three conceptual phases:

- **Phase I: Content Availability**
- **Phase II: Support Early Adopters**
- **Phase III: The Next Generation**



**Figure 18: Roadmap**

Each phase of the roadmap is divided into five sections (see Figure 19):

- A single statement of the overall **Objective** of this phase;
- An outline of the current **Situation** regarding TDM, in the context of this phase;
- A breakdown of the main **Challenges** to achieving this phase's objective;

<sup>7</sup> At FutureTDM [Knowledge Cafes](#) and other [community events](#)

<sup>8</sup> [D5.1 FutureTDM Policy Framework](#)

- **Key Principles** to address each challenge highlighted;
- **Activities** that the Commission and other parties should engage in to implement those principles and meet the phase's objective.

## PHASE I: CONTENT AVAILABILITY

Without data, there can be no data analytics. The first step towards increasing the use of TDM is to ensure that there is as much data available to practitioners as possible.

OBJECTIVE	ACTIVITIES
Ensure more large datasets are genuinely available to as many practitioners of TDM as possible.	<ol style="list-style-type: none"> <li>1 - Introduce harmonised, mandatory exceptions to copyright and database rights, for all TDM activities, without limitation to specific sectors or types of TDM practitioners.</li> <li>2 - Clarify any limitations to exceptions to intellectual property rights.</li> <li>3 - Clarify Data Protection frameworks to aid compliance.</li> <li>4 - Offer evaluation and certification of Data Protection practices to reduce uncertainty.</li> <li>5 - Commit to supporting open sharing of data in ways that are genuinely re-usable by machines as well as human readers.</li> <li>6 - Ensure funders and researchers make publicly-funded content genuinely open and re-usable for machines as well as human readers.</li> <li>7 - Support the development of centralised, integrated platforms providing access to data sources.</li> <li>8 - Encourage all content repositories to use open metadata standards and expose their metadata.</li> </ol>
SITUATION	
Without large datasets that are genuinely available for TDM – that is, legally and practically discoverable and re-usable – there can be no text or data mining. While large industries may have significant stores of private and proprietary content to work with, academic and smaller commercial practitioners can be severely limited by the lack of large, open datasets. The first step to increasing the uptake of TDM in Europe must be to increase the amount of data that is open for all TDM practitioners to use.	
CHALLENGES	
<ul style="list-style-type: none"> <li>Legal</li> <li>Knowledge/Awareness</li> <li>Technical</li> </ul>	
PRINCIPLES	
<ul style="list-style-type: none"> <li>Legal</li> <li>Knowledge/Awareness</li> <li>Technical</li> </ul>	
For more information see our <a href="#">Deliverable 5.4 Roadmap for increasing uptake of TDM</a>	

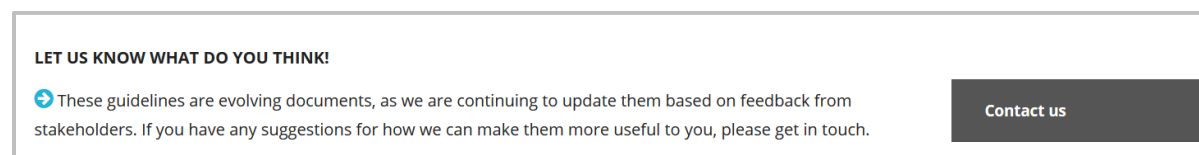
**Figure 19: Roadmap Phase I: Content Availability**

The challenges, principles and suggested activities outlined are all based on the extensive research and stakeholder engagement carried out over the course of the FutureTDM project – particularly the FutureTDM Policy Framework, which gives a more comprehensive overview of the ways in which all stakeholders in the TDM landscape can help to support the uptake of TDM technologies in Europe.

The Roadmap is available on the FutureTDM platform to increase the visibility and accessibility of these guidelines. The respective deliverable is also available online at the bottom of each roadmap webpage.

## 5 CONCLUSION

This stable and flexible FutureTDM platform wishes to empower practitioners with the resources, guidelines and solutions to promote the uptake of TDM in Europe. The platform was presented latest in the FutureTDM Symposium which took place within the iDSC Conference in Salzburg Austria on June 12<sup>th</sup> - 13<sup>th</sup> 2017 and received very positive comments. As indicated in almost all pages of the platform, feedback from stakeholders and any suggestions for improvement are welcome.



**Figure 20: Feedback requested on the FTDM platform**

The platform will continue to be revised and updated, as it has throughout the project, in order to maintain a coherent and up-to-date view on the European TDM landscape.